Institute of Analytics

The Global Body for Analytics

The Future is Here.

Dr Vinod on Text Mining 8971073111/9606945192
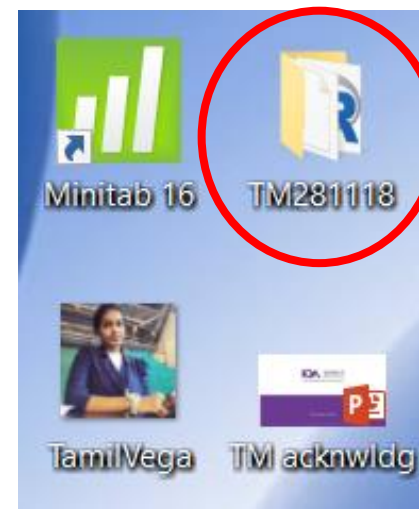vinod.murti@ioaglobal.org vinodanalytics@gmail.com

Set Working Directory
Keep document 'Acknowledgment' in TM190519

```
2  ########## READING THE DATA ##############
3
4  setwd("C:/Users/Dr Vinod/Desktop/TM281118")
5
```

See what is in 'Acknowledgment'

```
6  readLines("Acknowledgment.txt")
7
```

RE: In Appreciation of your contribution

Vinod [vinod@inurture.co.in]

Sent: Thu 8/24/2017 3:51 PM
To: 'Vasu SIR'
Cc: 'easwaran iyer'; 'R Lakshminarayanan'; 'Ashwin Ajilla'

Thank you very much for your kind words and acknowledgement, sir.
Actually, all those FDPs/MDPs/Workshops/Conferences were engineered perfectly because of your foresightedness and thought leadership! It was a perfect ensemble of Conceptual Understanding and Demonstration of Technical Skills. Moreover, not only me, but entire I Nurture team enjoy working with you because of your passionate leadership and a strong desire of positioning the university with remarkable academic edge.

Sir, it goes without saying that as a knowledge worker, I used to achieve tremendous satisfaction and happiness while doing such high end academic projects under your fabulous leadership and eagerly looking forward for such class assignments.

With Warm Regards,

Dr Vinod

Dear Dr. Vinod,

We place on record our heartfelt gratitude for your continuous efforts in designing the National Level FDP programs – *'Data Analysis through SPSS'*, in house FDPs on the same subject and FDPs on Excel for faculty members repeatedly. The recent National Level MDP on *'Business Analytics with R and Python'* was well received by faculty members and working professionals from the industry alike and has left us with positive enquires.

The ideas nurtured, guidance offered, papers reviewed during all the conferences, working as session chair during the conferences and the sessions conducted for teaching and learning communities - all have made a difference. Thank you for all the above and many more activities such as guiding faculty members through data analysis of their research work (Dr. Madhavi R, Ms. Smita S K, etc.), guest lectures for PG Commerce students, etc.

The new ideas designed include National Level MDP/FDPs on *'Data Visualization and Spreadsheet Modeling'*, MDP on *'Business Analytics with Python'*, etc. which will undoubtedly be the new feathers in the cap adorning the work of the University.

Research activities define a University's quality of work and all the above have contributed positively. We wish many more such successful programs and laurels to you in the years to come.

Regards,

Dr. B A Vasu

Center Head

School of Commerce-JU

Ph: +91-80-46501728/739

E-mail: vasu.ba@jainuniversity.ac.in

Website: www.jainuniversity.ac.in

```
> readLines("Acknowledgment.txt")
 [1] "Dear Dr. Vinod,"
```

18 Lines

```
 [2] "We place on record our heartfelt gratitude for your
continuous efforts in designing the National Level FDP pro
grams – 'Data Analysis through SPSS', in house FDPs on the
 same subject and FDPs on Excel for faculty members repeat
edly. The recent National Level MDP on 'Business Analytics
 with R and Python' was well received by faculty members a
nd working professionals from the industry alike and has l
eft us with positive enquires."
 [3] "The ideas nurtured, guidance offered, papers reviewe
d during all the conferences, working as session chair dur
ing the conferences and the sessions conducted  for teachi
ng and learning communities - all have made a difference.
Thank you for all the above and many more activities such
as guiding faculty members  through data analysis of their
research work (Dr. Madhavi R, Ms. Smita S K, etc.), guest
lectures for PG Commerce students, etc. "
```

```
[16] "Actually, all those FDPs/MDPs/Workshops/Conferences
were engineered perfectly because of your foresightedness
and thought leadership! It was a perfect ensemble of Conce
ptual Understanding and Demonstration of Technical Skills.
 Moreover, not only me, but entire I Nurture team enjoy wo
rking with you because of your passionate leadership and a
 strong desire of positioning the university with remarkab
le academic edge. "
[17] "Sir, it goes without saying that as a knowledge work
er, I used to achieve tremendous satisfaction and happines
s while doing such high end academic projects under your f
abulous leadership and eagerly looking forward for such cl
ass assignments. "


[18] "With Warm Regards,"
```

4

```
 8  # Let us have a vector of words and apply paste funtion into it
 9
10  a<- c("Hello", "World", "My", "First", "Handshake", "Through", "R")
11  a
12  # We are looking to use paste function and make it a chunk
13
14  b<- paste(a, collapse = " ")
15  b
16  # Just check how do these two look
17  a
18  b
```

Play around with

collapse = " "

paste
function

```
> a<- c("Hello", "World", "My", "First", "Handshake", "Through", "R")
> a
[1] "Hello"     "World"     "My"        "First"     "Handshake"
[6] "Through"   "R"
> b<- paste(a, collapse = " ")
> b
[1] "Hello World My First Handshake Through R"
> # Just check how do these two look
> a
[1] "Hello"     "World"     "My"        "First"     "Handshake"
[6] "Through"   "R"
> b
[1] "Hello World My First Handshake Through R"
```

The Future is Here

Lets club all 18 lines in ONE

```
26  chunk_2<- readLines("Acknowledgment.txt")
27
28  chunk_pasted_2 <- paste(chunk_2, collapse = " ")
29
30  #let us see how top of our data looks like
31  head(chunk_pasted_2)
```

```
> head(chunk_pasted_2)
[1] "Dear Dr. Vinod, We place on record our heartfelt gratitude for your con
tinuous efforts in designing the National Level FDP programs – 'Data Analysi
s through SPSS', in house FDPs on the same subject and FDPs on Excel for fac
ulty members repeatedly. The recent National Level MDP on 'Business Analytic
s with R and Python' was well received by faculty members and working profes
sionals from the industry alike and has left us with positive enquires. The
ideas nurtured, guidance offered, papers reviewed during all the conferences
, working as session chair during the conferences and the sessions conducted
  for teaching and learning communities - all have made a difference. Thank
you for all the above and many more activities such as guiding faculty membe
rs through data analysis of their research work (Dr. Madhavi R, Ms. Smita S
```

```
> readLines("Acknowledgment.txt")
 [1] "Dear Dr. Vinod,"

 [2] "We place on record our heartfelt gratitude for your
continuous efforts in designing the National Level FDP pro
grams – 'Data Analysis through SPSS', in house FDPs on the
 same subject and FDPs on Excel for faculty members repeat
edly. The recent National Level MDP on 'Business Analytics
 with R and Python' was well received by faculty members a
nd working professionals from the industry alike and has l
eft us with positive enquires."
 [3] "The ideas nurtured, guidance offered, papers reviewe
d during all the conferences, working as session chair dur
ing the conferences and the sessions conducted  for teachi
ng and learning communities - all have made a difference.
Thank you for all the above and many more activities such
as guiding faculty members through data analysis of their
research work (Dr. Madhavi R, Ms. Smita S K, etc.), guest
lectures for PG Commerce students, etc. "
```

```
[16] "Actually, all those FDPs/MDPs/Workshops/Conferences
were engineered perfectly because of your foresightedness
and thought leadership! It was a perfect ensemble of Conce
ptual Understanding and Demonstration of Technical Skills.
 Moreover, not only me, but entire I Nurture team enjoy wo
rking with you because of your passionate leadership and a
 strong desire of positioning the university with remarkab
le academic edge. "
[17] "Sir, it goes without saying that as a knowledge work
er, I used to achieve tremendous satisfaction and happines
s while doing such high end academic projects under your f
abulous leadership and eagerly looking forward for such cl
ass assignments. "

[18] "With Warm Regards,"
```

```
edge.  Sir, it goes without saying that as a knowledge worker, I used to ach
ieve tremendous satisfaction and happiness while doing such high end academi
c projects under your fabulous leadership and eagerly looking forward for su
ch class assignments.  With Warm Regards,"
>
```

Lets convert all to lower case

```
34  ########## CLEANING THE DATA ###########
35
36  # let us now lower case this data
37  clean_data1<- tolower(chunk_pasted_2)
38  head(clean_data1)
```

```
> head(chunk_pasted_2)
[1] "Dear Dr. Vinod, We place on record our heartfelt gratitude for your con
tinuous efforts in designing the National Level FDP programs – 'Data Analysi
s through SPSS', in house FDPs on the same subject and FDPs on Excel for fac
ulty members repeatedly. The recent National Level MDP on 'Business Analytic
s with R and Python' was well received by faculty members and working profes
sionals from the industry alike and has left us with positive enquires. The
ideas nurtured, guidance offered, papers reviewed during all the conferences
, working as session chair during the conferences and the sessions conducted
  for teaching and learning communities - all have made a difference. Thank
you for all the above and many more activities such as guiding faculty membe
rs through data analysis of their research work (Dr. Madhavi R, Ms. Smita S
```
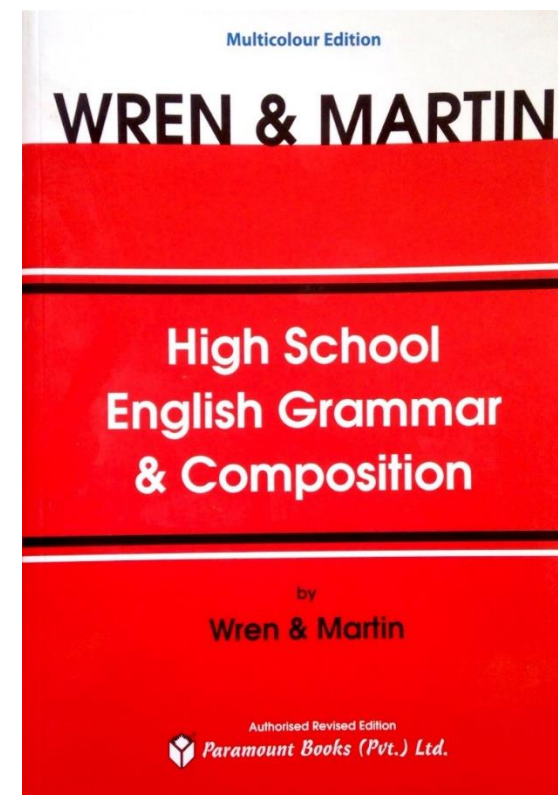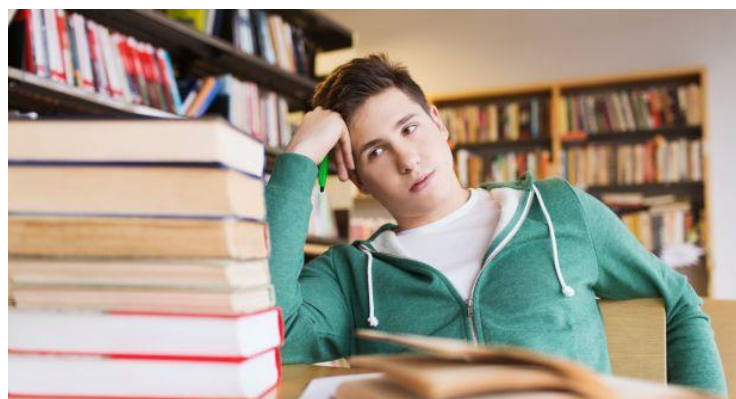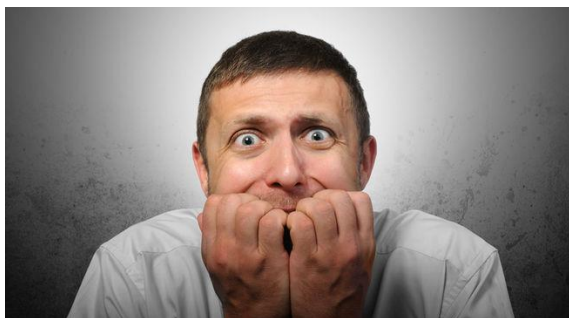
```
> clean_data1<- tolower(chunk_pasted_2)
> head(clean_data1)
[1] "dear dr. vinod, we place on record our heartfelt gratitude for your con
tinuous efforts in designing the national level fdp programs – 'data analysi
s through spss', in house fdps on the same subject and fdps on excel for fac
ulty members repeatedly. the recent national level mdp on 'business analytic
s with r and python' was well received by faculty members and working profes
sionals from the industry alike and has left us with positive enquires. the
ideas nurtured, guidance offered, papers reviewed during all the conferences
, working as session chair during the conferences and the sessions conducted
  for teaching and learning communities - all have made a difference. thank
```
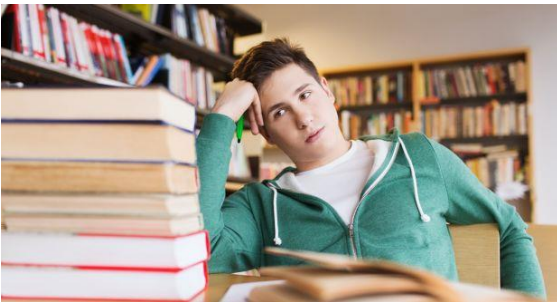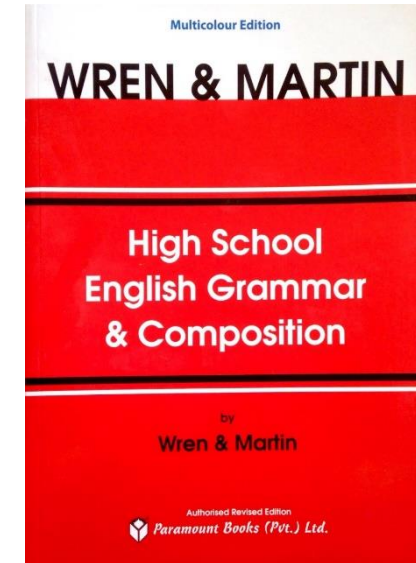
Why lower case?

Remove punctuations

```
40  # Cleaning the punctuations, pattern = "\\w"
41  # We are going to replace the punctuations by space,
42  # if we don't do so then we may make new words
43
44  clean_data2 <- gsub(pattern = "\\W", replace = " ",
45                 clean_data1)
46  head(clean_data2)
```

Remove punctuations


```
> clean_data1<- tolower(chunk_pasted_2)
> head(clean_data1)
[1] "dear dr. vinod, we place on record our heartfelt gratitude for your con
tinuous efforts in designing the national level fdp programs – 'data analysi
s through spss'  in house fdps on the same subject and fdps on excel for fac
ulty members repeatedly, the recent national level mdp on 'business analytic
s with r and python' was well received by faculty members and working profes
sionals from the industry alike and has left us with positive enquires. the
ideas nurtured, guidance offered, papers reviewed during all the conferences
, working as session chair during the conferences and the sessions conducted
 for teaching and learning communities – all have made a difference. thank
```

```
> head(clean_data2)
[1] "dear dr  vinod  we place on record our heartfelt gratitude for your con
tinuous efforts in designing the national level fdp programs  data analysi
s through spss  in house fdps on the same subject and fdps on excel for fac
ulty members repeatedly  the recent national level mdp on  business analytic
s with r and python  was well received by faculty members and working profes
sionals from the industry alike and has left us with positive enquires  the
ideas nurtured  guidance offered  papers reviewed during all the conferences
   working as session chair during the conferences and the sessions conducted
   for teaching and learning communities   all have made a difference  thank
```

Lets remove digits

```
48  # Cleaning the digits, pattern = "\\d"
49  # PROBABLY NOT REQUIRED!!
50
51  clean_data3 <- gsub(pattern = "\\d", replace = " ",
52                      clean_data2)
53  head(clean_data3)
```

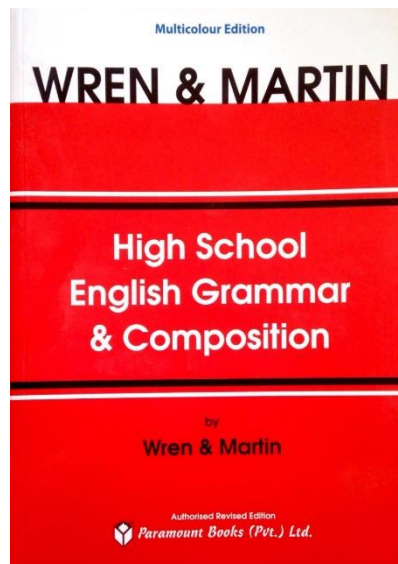**Digits have no relevance in Text Analysis**

```
> clean_data3 <- gsub(pattern = "\\d", replace = " ",
+                     clean_data2)
> head(clean_data3)
[1] "dear dr  vinod  we place on record our heartfelt gratitude for your con
tinuous efforts in designing the national level fdp programs    data analysi
s through spss   in house fdps on the same subject and fdps on excel for fac
ulty members repeatedly  the recent national level mdp on  business analytic
s with r and python  was well received by faculty members and working profes
```

The Future is Here

Stop Words

```
55  # Cleaning the stopwords
56  install.packages('tm')
57  library('tm')
58
59  #let see a preview of stopwords
60  stopwords()
```

```
> #let see a preview of stopwords
> stopwords()
  [1] "i"          "me"         "my"           "myself"       "we"
  [6] "our"        "ours"       "ourselves"    "you"          "your"
 [11] "yours"      "yourself"   "yourselves"   "he"           "him"
 [16] "his"        "himself"    "she"          "her"          "hers"
 [21] "herself"    "it"         "its"          "itself"       "they"
 [26] "them"       "their"      "theirs"       "themselves"   "what"
 [31] "which"      "who"        "whom"         "this"         "that"
 [36] "these"      "those"      "am"           "is"           "are"
 [41] "was"        "were"       "be"           "been"         "being"
 [46] "have"       "has"        "had"          "having"       "do"
 [51] "does"       "did"        "doing"        "would"        "should"
 [56] "could"      "ought"      "i'm"          "you're"       "he's"
 [61] "she's"      "it's"       "we're"        "they're"      "i've"
 [66] "you've"     "we've"      "they've"      "i'd"          "you'd"
 [71] "he'd"       "she'd"      "we'd"         "they'd"       "i'll"
 [76] "you'll"     "he'll"      "she'll"       "we'll"        "they'll"
 [81] "isn't"      "aren't"     "wasn't"       "weren't"      "hasn't"
 [86] "haven't"    "hadn't"     "doesn't"      "don't"        "didn't"
 [91] "won't"      "wouldn't"   "shan't"       "shouldn't"    "can't"
 [96] "cannot"     "couldn't"   "mustn't"      "let's"        "that's"
[101] "who's"      "what's"     "here's"       "there's"      "when's"
[161] "most"       "other"      "some"         "such"         "no"
[166] "nor"        "not"        "only"         "own"          "same"
[171] "so"         "than"       "too"          "very"
```
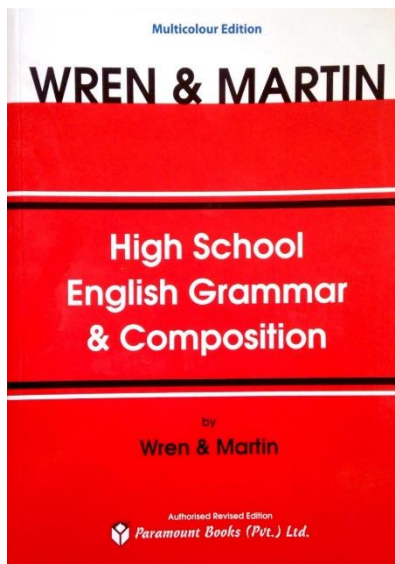
Remove Stop Words

```
62  #let us remove them using function removeWords()
63  clean_data4 <- removeWords(clean_data3,
64                              stopwords())
65  head(clean_data4)
```

Which STOP WORDS gone?

```
> head(clean_data4)
[1] "dear dr  vinod    place   record  heartfelt gratitude    continuous effort
s  designing  national  level fdp programs     data analysis  spss     house fd
ps     subject  fdps  excel  faculty members repeatedly    recent national lev
el mdp    business analytics   r  python    well received  faculty members   wor
king professionals    industry alike   left us  positive enquires    ideas nur
tured  guidance offered  papers reviewed    conferences   working  session ch
air   conferences   sessions conducted   teaching  learning communities
made  difference  thank     many  activities   guiding faculty members  da
ta analysis    research work  dr  madhavi r  ms  smita s k  etc    guest lect
ures  pg commerce students  etc    new ideas designed include national level
 mdp fdps    data visualization  spreadsheet modeling    mdp   business analyt
ics  python   etc   will undoubtedly   new feathers   cap adorning   work  u
niversity  research activities define  university s quality  work     contr
ibuted positively   wish many   successful programs  laurels    years   come
   regards   dr  b  vasu center head school   commerce ju ph
   e mail  vasu ba jainuniversity ac   website  www jainuniversity ac    resp
ected dr vasu    thank    much   kind words  acknowledgement  sir   actually
```

Multicolour Edition

WREN & MARTIN

High School English Grammar & Composition

by Wren & Martin

Authorised Revised Edition
Paramount Books (Pvt.) Ltd.

Dr Vinod on Text Mining 8971073111/9606945192
vinod.murti@ioaglobal.org vinodanalytics@gmail.com

The Future is Here

> head(clean_data2)
[1] "dear dr vinod we place on record our heartfelt gratitude for your con
tinuous efforts in designing the national level fdp programs ▮ data analysi
s through spss ▮ in house fdps on the same subject and fdps on excel for fac
ulty members repeatedly the recent national level mdp on business analytic
s with r and python was well received by faculty members and working profes
sionals from the industry alike and has left us with positive enquires the
ideas nurtured guidance offered papers reviewed during all the conferences
  working as session chair during the conferences and the sessions conducted
  for teaching and learning communities   all have made a difference   thank

> #let see a preview of stopwords
> stopwords()
     [1] "i"   "me"   "my"   "myself"   "we"
     [6] "our"   "ours"   "ourselves"   "you"   "your"
    [11] "yours"   "yourself"   "yourselves" "he"   "him"
    [16] "his"   "himself"   "she"   "her"   "hers"
    [21] "herself"   "it"   "its"   "itself"   "they"
    [26] "them"   "their"   "theirs"   "themselves" "what"
    [31] "which"   "who"   "whom"   "this"   "that"
    [36] "these"   "those"   "am"   "is"   "are"
    [41] "was"   "were"   "be"   "been"   "being"
    [46] "have"   "has"   "had"   "having"   "do"
    [51] "does"   "did"   "doing"   "would"   "should"
    [56] "could"   "ought"   "i'm"   "you're"   "he's"
    [61] "she's"   "it's"   "we're"   "they're"   "i've"
    [66] "you've"   "we've"   "they've"   "i'd"   "you'd"
    [71] "he'd"   "she'd"   "we'd"   "they'd"   "i'll"
    [76] "you'll"   "he'll"   "she'll"   "we'll"   "they'll"
    [81] "isn't"   "aren't"   "wasn't"   "weren't"   "hasn't"
    [86] "haven't"   "hadn't"   "doesn't"   "don't"   "didn't"
    [91] "won't"   "wouldn't"   "shan't"   "shouldn't"   "can't"
    [96] "cannot"   "couldn't"   "mustn't"   "let's"   "that's"
   [101] "who's"   "what's"   "here's"   "there's"   "when's"
   [161] "most"   "other"   "some"   "such"   "no"
   [166] "nor"   "not"   "only"   "own"   "same"
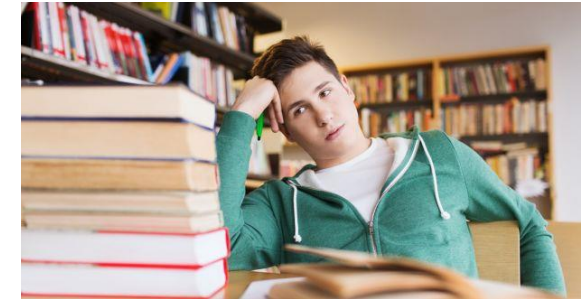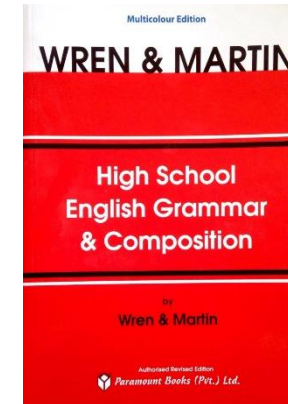   [171] "so"   "than"   "too"   "very"
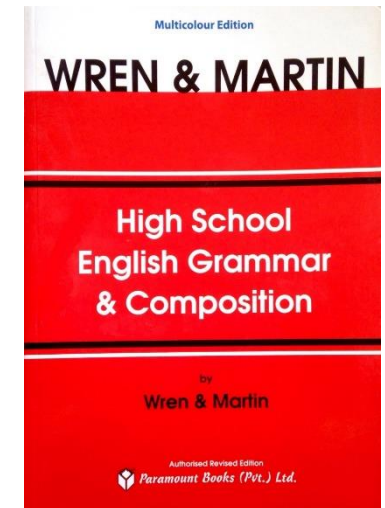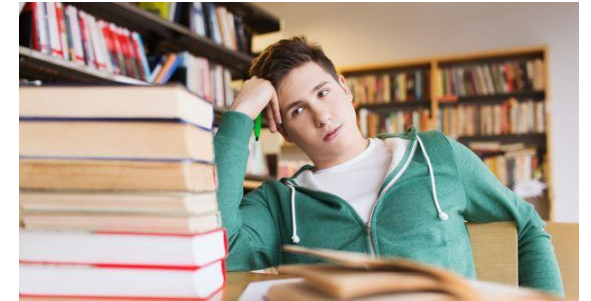
Note the size reduction!

> head(clean_data4)
[1] "dear dr vinod   place record heartfelt gratitude   continuous effort
s designing national level fdp programs   data analysis spss   house fd
ps   subject fdps excel faculty members repeatedly   recent national lev
el mdp   business analytics r python   well received faculty members wor
king professionals   industry alike   left us positive enquires   ideas nur
tured guidance offered papers reviewed   conferences   working session ch
air   conferences   sessions conducted   teaching learning communities
made difference thank   many activities   guiding faculty members da
ta analysis   research work dr madhavi r ms smita s k etc   guest lect
ures pg commerce students etc   new ideas designed include national level
 mdp fdps   data visualization spreadsheet modeling   mdp   business analyt
ics python etc   will undoubtedly   new feathers   cap adorning work   u
niversity research activities define university s quality work   contr
ibuted positively   wish many   successful programs laurels   years come
  regards dr b vasu center head school   commerce ju ph
  e mail vasu ba jainuniversity ac website www jainuniversity ac   resp
ected dr vasu   thank   much   kind words acknowledgement sir   actually

The Future is Here

Single letters

```
67  # let us remove single letters, here \\b[A-z]
68  # represents that string starts with any letter
69  # between a-z
70  # and string can take uppercase letters as well
71  # as lower case letters
72  # and the subsequent \\b{1} says that the
73  # string ends with length one
74  clean_data5 <- gsub(pattern = "\\b[A-z]\\b{1}",
75                      replace = " ", clean_data4)
76  clean_data5
77
```

Single letters

```
> head(clean_data4)
[1] "dear dr   vinod    place   record   heartfelt gratitude    continuous effort
s  designing   national  level fdp programs      data analysis  spss     house fd
ps    subject  fdps  excel  faculty members repeatedly    recent national lev
el mdp   business analytics  [r]  python   well received   faculty members   wor
king professionals    industry alike    left us  positive enquires    ideas nur
tured  guidance offered  papers reviewed    conferences   working   session ch
air   conferences    sessions conducted    teaching   learning communities
made   difference  thank      many   activities    guiding faculty members   da
ta analysis    research work   dr   madhavi  [r]  ms  smita  [s k]  etc     guest lect
ures   pg commerce students   etc     new ideas designed include national level
 mdp fdps    data visualization  spreadsheet modeling    mdp    business analyt
ics python    etc   will undoubtedly   new feathers   cap adorning   work    u
niversity   research activities define   university  s quality   work      contr
ibuted positively   wish many    successful programs  laurels     years   come
   regards   dr  [b]  vasu center head school   commerce ju ph
  [e] mail  vasu ba jainuniversity ac   website   www jainuniversity ac     resp
ected dr vasu    thank    much   kind words   acknowledgement  sir    actually
```



```
> clean_data5
[1] "dear dr   vinod    place   record   heartfelt gratitude    continuous efforts   designing
 national level fdp programs      data analysis  spss     house fdps      subject  fdps  excel
  faculty members repeatedly    recent national level mdp    business analytics      python
   well received   faculty members   working professionals    industry alike    left us  posit
ive enquires    ideas nurtured  guidance offered  papers reviewed    conferences   working
  session chair   conferences    sessions conducted    teaching   learning communities      ma
de   difference  thank      many   activities    guiding faculty members   data analysis    r
esearch work   dr   madhavi    ms  smita      etc    guest lectures   pg commerce students
etc     new ideas designed include national level mdp fdps    data visualization   spreadshe
et modeling    mdp    business analytics  python    etc   will undoubtedly   new feathers
cap adorning   work    university   research activities define   university   quality   work
    contributed positively   wish many    successful programs  laurels     years   come  re
gards   dr     vasu center head school   commerce ju ph                 mail   vasu b
a jainuniversity ac   website   www jainuniversity ac     respected dr vasu    thank    much
  kind words   acknowledgement  sir    actually    fdps mdps workshops conferences   engineer
ed perfectly    foresightedness   thought leadership    perfect ensemble   conceptual unde
rstanding   demonstration  technical skills  moreover      entire   nurture team enjoy wor
king        passionate leadership    strong desire  positioning  university    remarkable acad
```
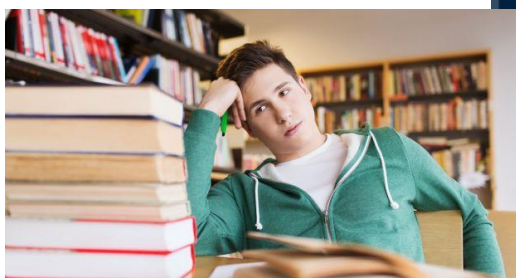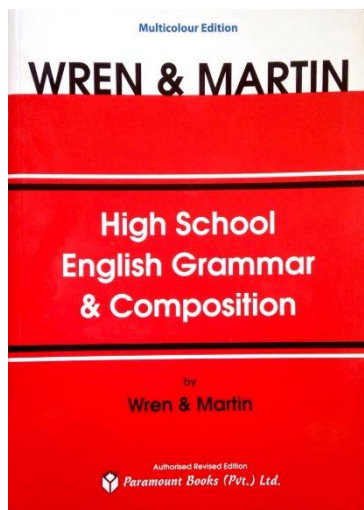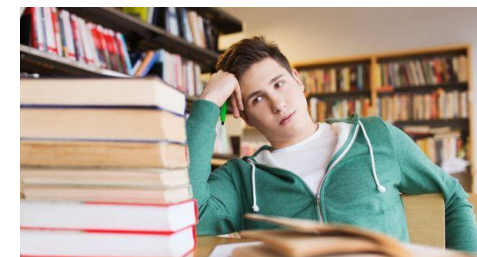
Dr Vinod on Text Mining 8971073111/9606945192
vinod.murti@ioaglobal.org vinodanalytics@gmail.com

The Future is Here

Remove White Spaces

```
79  # We can finally remove whitespaces using
80  # stripWhitespace() function, which is part of tm library
81  clean_data6 <- stripWhitespace(clean_data5)
82  head(clean_data6)
```

```
> clean_data6 <- stripWhitespace(clean_data5)
> head(clean_data6)
[1] "dear dr vinod place record heartfelt gratitude continuous effor
ts designing national level fdp programs data analysis spss house fd
ps subject fdps excel faculty members repeatedly recent national lev
el mdp business analytics python well received faculty members worki
ng professionals industry alike left us positive enquires ideas nurt
ured guidance offered papers reviewed conferences working session ch
air conferences sessions conducted teaching learning communities mad
e difference thank many activities guiding faculty members data anal
ysis research work dr madhavi ms smita etc guest lectures pg commerc
```

Dr Vinod on Text Mining 8971073111/9606945192
vinod.murti@ioaglobal.org vinodanalytics@gmail.com

The Future is Here

Split words

```
84  ###### FREQUENCY OF THE WORDS  ############
85  # We now have a chunk of lines, and we are
86  # looking for counting the words
87  # If you remember we had joined variouslines
88  # and made a chunk
89  # So we split individual words and add a space
90  # between them as splitter
91  clean_data7<- strsplit(clean_data6," ")
92  head(clean_data7)
```

```
> clean_data7<- strsplit(clean_data6," ")
> head(clean_data7)
[[1]]
  [1] "dear"         "dr"           "vinod"
  [4] "place"        "record"       "heartfelt"
  [7] "gratitude"    "continuous"   "efforts"
 [10] "designing"    "national"     "level"
 [13] "fdp"          "programs"     "data"
 [16] "analysis"     "spss"         "house"
 [19] "fdps"         "subject"      "fdps"
 [22] "excel"        "faculty"      "members"

[202] "forward"      "class"        "assignments"
[205] "warm"         "regards"
```

206 words

Dr Vinod on Text Mining 8971073111/9606945192
vinod.murti@ioaglobal.org vinodanalytics@gmail.com

17

The Future is Here

Frequency of words


```
 95  word_freq1<-table(clean_data7)
 96  head(word_freq1)
 97
 98  word_freq2<- cbind(names(word_freq1),
 99               as.integer(word_freq1))
100  head(word_freq2)
```

```
> word_freq1<-table(clean_data7)
> head(word_freq1)
clean_data7
        ac        academic          achieve acknowledgement
         2               2                1               1
 activities        actually
          2               1
```

```
> word_freq2<- cbind(names(word_freq1),
+                 as.integer(word_freq1))
> head(word_freq2)
      [,1]                  [,2]
[1,]  "ac"                  "2"
[2,]  "academic"            "2"
[3,]  "achieve"             "1"
[4,]  "acknowledgement"     "1"
[5,]  "activities"          "2"
[6,]  "actually"            "1"
```
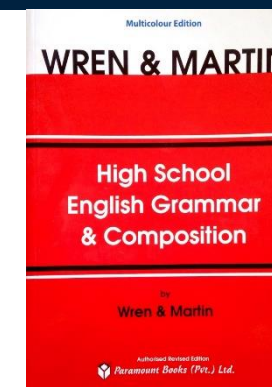
Dr Vinod on Text Mining 8971073111/9606945192
vinod.murti@ioaglobal.org vinodanalytics@gmail.com

The Future is Here
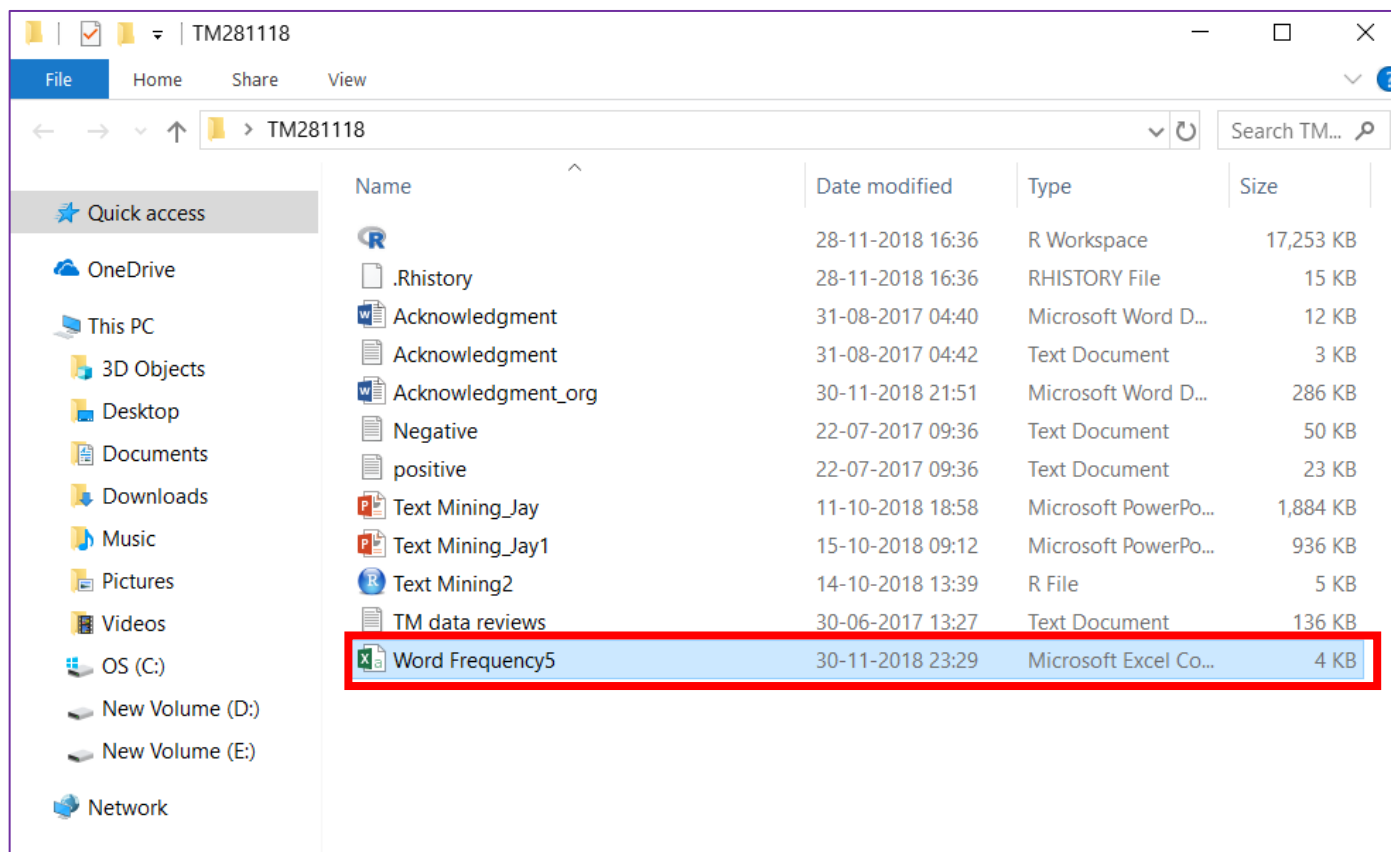
Word Clouds

```
104  ###### Word Cloud ########
105  # we will use word cloud library to come up with word cloud
106  install.packages("RColorBrewer")
107  install.packages("wordcloud")
108
109  library(RColorBrewer)
110  library(wordcloud)
```
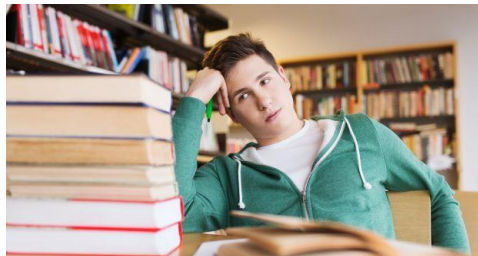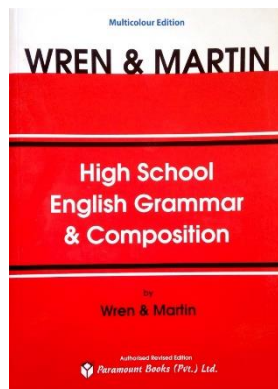
```
112  # After installing packages,
113  # we want to organize our words as per wordcloud()
114
115  class(clean_data7) #Class of data stored as words
116  # We want the class to be as characters
117  # One way to do is to unlist the list
118  word_cloud1<-unlist(clean_data7)
119
```



```
> class(clean_data7) #Class of data stored as words
[1] "list"
> # We want the class to be as characters
> # One way to do is to unlist the list
> word_cloud1<-unlist(clean_data7)
```

The Future is Here

```
120 wordcloud(word_cloud1) # The most raw form of wordcloud
121 wordcloud(word_cloud1, min.freq = 3)
122 #A slightly better version
```

Dr Vinod on Text Mining 8971073111/9606945192
vinod.murti@ioaglobal.org vinodanalytics@gmail.com

```
123  wordcloud(word_cloud1, min.freq = 1,
124           random.order = FALSE)
```

Try with
random.order = TRUE

The Future is Here
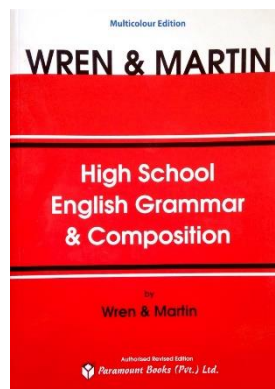
```
126  # We want to center the major words
127  # We want frequent words to be 4 times larger
128  # in shape than infrequent words
129  wordcloud(word_cloud1, min.freq = 3,
130          random.order = FALSE, scale = c(4,1))
131
```
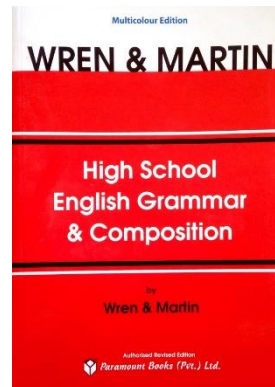
```
132  # We want to add colors to words,
133  # we chose rainbow function to add multiple colors,
134  # number of clors in parentheses
135
136  pal2 <- brewer.pal(5, "Dark2")
137  wordcloud(word_cloud1, min.freq = 3,
138           random.order = FALSE, scale = c(4,1),
139           colors = pal2)
140
141  ??brewer.pal
```

Encourage you to learn more about **brewer.pal**

Multicolour Edition
WREN & MARTIN

High School English Grammar & Composition

by
Wren & Martin

Authorised Revised Edition
Paramount Books (Pvt.) Ltd.

vasu
university
conferences
data etc
national
work fdps
mdp
level
faculty members
working
leadership

Dr Vinod on Text Mining 8971073111/9606945192
vinod.murti@ioaglobal.org vinodanalytics@gmail.com

The Future is Here
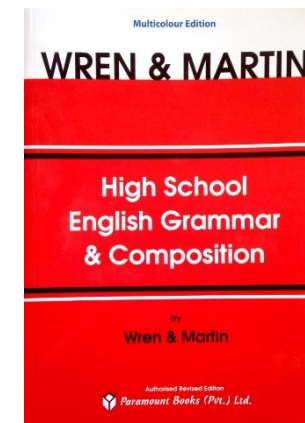
Sentiment Analysis
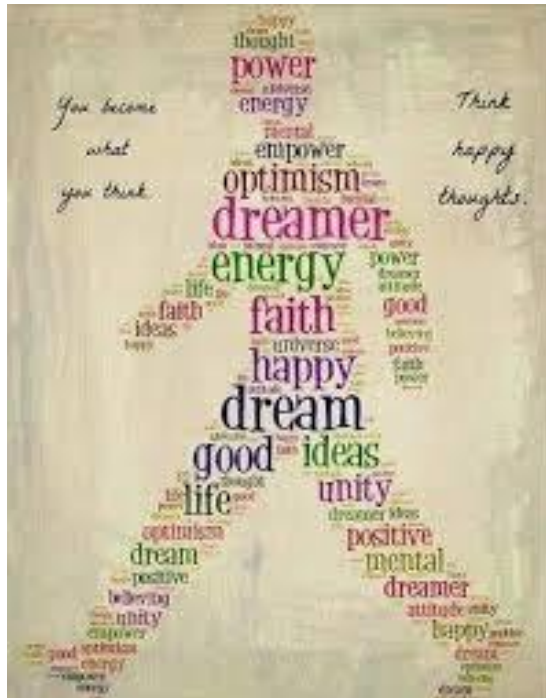
```
143 ##### Sentiment Analysis  ######
144 # Getting the bags of positive and negative words
145 positive<- scan("positive.txt",
146                 what= "character",
147                 comment.char = ";")
148
149 negative<- scan("negative.txt",
150                 what= "character",
151                 comment.char = ";")
152
153 head(positive)
154 head(negative)
155
```

```
> positive<- scan("positive.txt",
+                 what= "character",
+                 comment.char = ";")
Read 2007 items
> negative<- scan("negative.txt",
+                 what= "character",
+                 comment.char = ";")
Read 4784 items
> head(positive)
[1] "ï»¿"        "a+"         "abound"     "abounds"    "abundance"
[6] "abundant"
> head(negative)
[1] "ï»¿"        "2-faced"    "2-faces"    "abnormal"
[5] "abolish"    "abominable"
```

WALK AWAY
FROM ANYTHING
THAT GIVES YOU
BAD VIBES. THERE
IS NO NEED TO
EXPLAIN OR MAKE
SENSE OF IT. IT'S
YOUR LIFE.
DO WHAT MAKES
YOU HAPPY.
THEGOODVIBE.CO

Sentiment Analysis
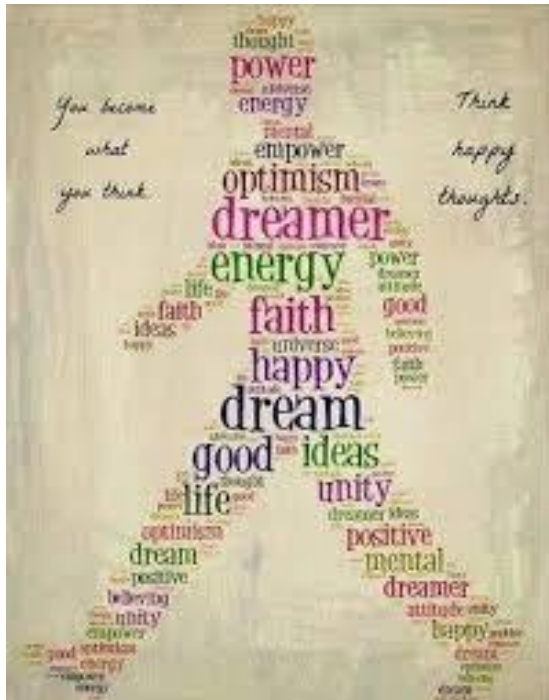
```
156  # Now we have our positive and negative words
157  # We will use them to match words in our text,
158  # using match()
159  senti_analysis<- unlist(clean_data7)
160
161  match(senti_analysis, positive)
```

```
> senti_analysis<- unlist(clean_data7)
> match(senti_analysis, positive)
  [1]    NA    NA    NA    NA    NA   900   857    NA    NA    NA    NA    NA
 [13]    NA    NA    NA    NA    NA    NA    NA    NA    NA   620    NA    NA
 [25]    NA    NA    NA    NA    NA    NA    NA    NA  1937    NA    NA    NA
 [37]    NA    NA    NA    NA    NA    NA  1323    NA    NA    NA   864    NA
 [49]    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA
 [61]    NA    NA  1793    NA    NA    NA    NA    NA    NA    NA    NA  1988
 [73]    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA
```

Dr Vinod on Text Mining 8971073111/9606945192
vinod.murti@ioaglobal.org vinodanalytics@gmail.com

The Future is Here

Sentiment Analysis

```
163  match(senti_analysis, negative)
164
```

```
> match(senti_analysis, negative)
  [1] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
 [21] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
 [41] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
 [61] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
 [81] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
[101] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
[121] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
[141] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
[161] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
[181] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
[201] NA NA NA NA NA NA
```

NO NEGATIVE THOUGHTS ALLOWED

shutterstock.com · 1099939772

```
176  p_score <- sum(!is.na(match(senti_analysis,
177                                positive)))
178  p_score
179  n_score <- sum(!is.na(match(senti_analysis,
180                                negative)))
181  n_score
182
183  Sentiment_score = p_score - n_score
184  Sentiment_score
```

NO NEGATIVE THOUGHTS ALLOWED
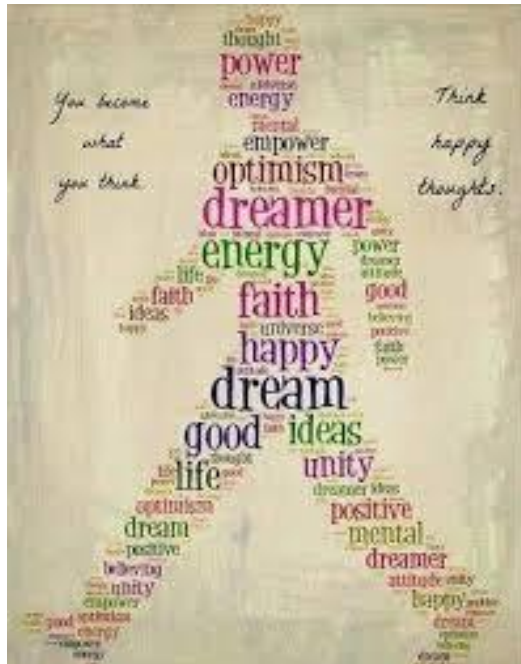

```
> p_score <- sum(!is.na(match(senti_analysis,
+                              positive)))
> p_score
[1] 23
> n_score <- sum(!is.na(match(senti_analysis,
+                              negative)))
> n_score
[1] 0
> Sentiment_score = p_score - n_score
> Sentiment_score
[1] 23
```

Dr Vinod on Text Mining 8971073111/9606945192
vinod.murti@ioaglobal.org vinodanalytics@gmail.com

29

The Future is Here